DOCUMENT RESUME

ED 086 228 IR 000 080

AUTHOR	Abbott, George L.
TITLE	Indexing for the Growing Instructional Media Center.
INSTITUTION	Stanford Univ., Calif. ERIC Clearinghouse on Educational Media and Technology.
SPONS AGENCY	National Inst. of Education (DHEW), Washington, D.C.
PUB DATE	Sep 73
CONTRACT	NE-C-00-4-0027
NOTE	25p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Classification; Filing; *Indexing; *Instructional Materials Centers; Instructional Media; Library Science; State of the Art Reviews
IDENTIFIERS	Library of Congress; Sears List of Subject Headings

ABSTRACT

Since indexing systems concentrate upon the information content of materials and not upon their form, instructional media centers (IMC) can use one system for all media. Content descriptors can be selected from a thesaurus of accepted terms, from the title of the material, or from an analysis of the content. The first of these three methods is the most satisfactory for dealing with multiple forms of media; the Sears List of Subject Headings and Subject Headings used in the Dictionary Catalogs of the Library of Congress are the most commonly used thesauri. It is recommended that the main file index of the IMC contain all entries for all materials and that in-depth indexing be provided through use of several descriptors for each item. Lastly, catalog card files should be employed wherever possible. (PB)

FILMED FROM BEST AVAILABLE COPY

ED 086228

INDEXING FOR THE GROWING INSTRUCTIONAL MEDIA CENTER

George L. Abbott

September 1973

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

IR 000 080

TABLE OF CONTENTS

Introduction	1
Media as Information	2
Definition of Indexing	4
Overview of Indexing Systems	6
Applicability of Existing Indexing Systems to Media	13
Summary	16
Bibliography	17
Selected List of Thesauri and Media Indexes	21

INTRODUCTION

Instructional Media Centers are being confronted with increasing collections of media in many new formats. It sometimes appears as though a new media format or some variation of an old format is introduced every week. These new formats raise many questions on their handling in the Instructional Media Center. The specific question we will treat here is: how to index these media for ease of retrieval by users.

MEDIA AS INFORMATION

All media exist for the communication of information or ideas. The media, of and by itself, is not the content of this information but merely a carrier - a road on which to transport the information.

Communication models have been set up by many educators to explore the ways information is transmitted from one individual to another. The simplest model consists of three elements.

The example, listening to an audio tape can be used to explain these elements. You, the listener, are the receiver; the media, audio tape, is the channel; and the words or information conveyed originate from the sender. The tape itself does not in any way alter the factual information presented and that information could also be conveyed by printed or other means. Although the information or ideas conveyed are independent of the media, the media may in some cases allow for giving interpretation to the information, by voice expression for example. This in no way alters the factual base of the information.

Since indexing systems concentrate on the information content of materials, those systems used for printed materials can work equally as well for films, or tapes, or slides, or any form of media. In all but the very largest collections of materials general indexing systems

will suffice for all media. In extremely large or comprehensive subject media collections, special indexing systems may need to be developed, but their development would parallel the present development of special indexing systems and classification schemes for collections of medical, legal and architectural materials. These special indexing systems, such as MeSH (Medical Subject Headings), are applicable to all media.

DEFINITION OF INDEXING

According to Ms. Hilda Feinberg in her book Title Derivative Indexing Techniques:

> Indexing consists of indicating the subject content of an item of information by assigning one or more terms to the document so as to characterize it. The word "term" is used broadly to include any form of class, subclass, subject heading, uniterm, compound word or phrase.

Indexing, therefore, is that specific activity of identifying the subject or informational content of an item. This is in contrast to the general concept of cataloging where descriptive information is the chief attribute. There are countless cataloging manuals for non-print media. Many states, school districts and associations have published one. These cataloging manuals outline in detail how to describe a physical item including the appropriate media designator, lengths, speeds, sizes and other identifying factors. For the most part these manuals concern themselves with the physical attributes of the item and not the informational or subject content. Very little is said in most of these manuals about subject content. Usually a one sentence statement is made relating to the subject content such as "use Sears subject headings" or "classify in Dewey".

> These general statements may seem insufficient and lacking in detail, but cataloging manuals concentrate on access by physical form of media, rather than access by subject or content.

A concept frequently used in indexing is "entry point." This concept can best be explained by examining a user search. When a user chooses a term and looks for that term in a completed index, that term becomes the "entry point' or "access point"--the door into the index to secure the information. The terms or words used to look up information are entry points.

While the main topic discussed here is media indexing from a subject point of view, indexing on grade level, date of production, or other significant information can be handled through an added entry process, to be discussed later.

OVERVIEW OF INDEXING SYSTEMS

The number of existing indexing systems extends in the hundreds: some are general information systems, most are highly specific. The basic difference in these systems usually is the list of terms and the method for arriving at these terms. For highly specialized and exhaustive collections in a given subject field special thesauri or word lists are constructed to allow for retrieval of documents at an acceptable level. In any indexing system the acceptable level in retrieving documents is that point where the maximum number of documents is retrieved with the minimum number of documents inappropriate to your request. All indexing systems are based on assigning a word or words to a subject content of an item or in the case of a classified index, assigning a code or classification number which can be converted to subject terms. If an item consists of multiple concepts, more than one term would be assigned.

Determination of the retrieval term can be from three sources. One is the thesaurus, or approved list of terms to be used in describing items in a given collection. This thesaurus method includes classification lists where numbers are used as terms. A second method is the KWIC (Keyword in context) or KWOC (Keyword out of context) index where the terms consist of words taken directly from the title of the item. The third, a more sophisticated, mechanical system is automatic indexing where the terms are derived from the contents of the document by analyzing the frequency of occurence of words in the text. These words are then used as descriptors.

The thesaurus is the most common and simplest form of indexing for small to medium general collections. By this method, terms are assigned from a preprescribed list of terms to indicate the subject content of the item. This master list of terms or thesaurus is constantly being updated to include new terms and relationships. Cross references from similar terms are usually included. The advantage of the thesaurus approach is the existence of a definite list of terms. The student is able to refer to a finite list of entry points for his search. He must, however, translate his request into terms from this list.

Classification is related to thesaurus indexing. The classification scheme is basically a thesaurus with numbers or letters assigned to each term. The class number then becomes a compound entry representing many terms related to the contents of the item. Since classification schemes combine many terms in one class number, they provide a single entry point usually under the most prevalent subject. Additional entry points are generally not available.

Indexing by the KWIC or KWOC method is based solely on the title of the item. The index is prepared by providing entry points under each significant word in the title. The number of entries is determined by the number of significant words in the title. To do an exhaustive search on any topic all synonyms would need to be checked since the documents are not grouped under specific terms. There are no cross references in a KWIC/KWOC index. An advantage to this method is no intermediate step of referring to a thesaurus is necessary. The index

is in a natural language and the terms used are those used by the author. This index can most easily be prepared mechanically by computer by inputing the titles of all documents to be indexed. As an example, a KWIC index for the two titles "An Inquiry into the Uses of Instructional Technology" and "Educational Technology in the Seventies" is as follows.

```
seventies                    EDUCATIONAL technology in the
technology              An INQUIRY into the uses of instructional
n inquiry into the uses of INSTRUCTIONAL technology                A
ucationaltechnology in the SEVENTIES                              Ed
                 Educational TECHNOLOGY in the seventies
the uses of instructional  TECHNOLOGY            An inquiry into
       An inquiry into the USES of instructional technology
```

As mentioned above, KWIC indexes are not cross-referenced. Therefore, one would have to check related terms, such as audio-visual media, etc, to locate additional references, and both instructional technology and educational technology would need to be used to retrieve information.

A KWOC index is constructed similarly to a KWIC index except the keyword is repeated out of context usually in a column at the beginning of each line. The first two entries from our previous example would appear as follows.

```
EDUCATIONAL    Educational technonlogy in the seventies
INQUIRY        An inquiry into the uses of instructional technology
```

Another automated but more sophisticated method is automatic indexing. This method also uses the author's terms directly but the entire document is used in determining the index terms. Terms are used as entry points if they occur with a predetermined frequency in the document. The disadvantage of this method is the entire document must be entered into the

computer. Optical scanning and other technological improvements may aid in this input.

In addition to the variance in indexing methods for determining entry points, differences also occur in the methods of file organization. Once the word list has been established there are several methods of file organization for storing the completed index. Two of the most common types are the catalog card file and the printed book catalog. Other special methods include edge notch cards, peek-a-boo or optical coincidence and computer data bases. These latter systems are most helpful for coordinate indexing- that is locating documents with at least two specified concepts.

In a card file, entries are arranged alphabethically under the subject terms and sufficient information to be able to locate the item is also given. This file method is usable when the indexing method is a thesaurus or classification list. One advantage of the card file is it is easy to update.

The printed list or book catalog is useful for all indexing methods and can vary from an exact copy of an existing card file to a computer generated KWIC index or automatic indexing output. An advantage to the printed catalog is its portability. It is, however, difficult to keep current.

In an edge notch system cards are used with holes around the edge. These holes are assigned subject meaning and each card represents a document. The holes representing the subject contents of a given document are notched out. In this method, a thin rod (is slid) through

the specific hole for the subject desired in the stack of document cards. Those containing that subject will drop out of the file. Coordinate indexing can be accomplished by the use of two or more rods. In that instance only those documents with all the subjects sought will drop.

The peek-a-boo or optical coincidence method is similar to the edge notch method. With peek-a-boo indexing each card is representative of a subject. On each card is a grid of small boxes, each box representing a document. These boxes are punched out for all documents containing a specific subject on that subject card. To locate documents on two or more subjects, the card for each subject is pulled from the master file. The cards are held together to the light and any area with a hole through all the cards chosen represents a document containing those subjects.

Another related system is uniterm or terminal digit indexing. Again in this system each subject is represented by a card. These cards are divided into 10 columns and the serial numbers for documents on a specified subject are entered on that subject card in columns according to the last digit of their serial number. This method makes visual scanning of the subject cards for document retrieval easier.

Both of these systems break down with collections in excess of 10,000 items. Their main advantage is the ability to do coordinate searching and thereby increase the precision on searches.

Because of the extreme spped and large storage capabilities of the computer it is able to handle any of the indexing methods described. Edge notch and peek-a-boo are manual systems recreating some of the combinatory possibilities of a computer. The only method thus far described for both indexing and file organization that the computer cannot completely handle is the thesaurus approach. With the thesaurus an indexer must chose the term or terms appropriate to describe the document. From that point the computer canthen process this information, In addition to being able to operate under any of the

above mentioned systems and to give output in the forms prescribed by these systems, the conputer is also capable of a file organization of its own and will retrieve document references directly from an internal master file upon request.

The number of entry points or terms used to index a document is important in its retrievability. By assigning multiple entries, as many as are felt necessary, all important subjects represented in an item can be indexed, and the item is more easily retrieved. This is especially necessary with the multidisciplinary nature of materials today and a single subject is no longer sufficient to adequately index an item. Most indexing methods allow for multiple entry points. As we have said, however, classification schemes can provide only a single entry point.

APPLICABILITY OF EXISTING INDEXING SYSTEMS TO MEDIA

Having reviewed the state of indexing, how does the individual in an Instructional Media Center with a starting collection determine the indexing method to be used? What method of file organization is best? Although subject content is independent of form we can see some indexing methods are dependent to an extent on printed word and are not applicable to purely visual materials. In so far as possible each Instructional Media Center should adopt one indexing method for all media - print and non-print.

Of the methods of indexing, the thesaurus or subject words approach is the most satisfactory for dealing with media. Since media titles are often less meaningful than book titles KWIC indexing poses a problem. Automatic indexing would require word input and as stated some media is purely visual.

The two most commonly used thesauri in libraries are Sears List of Subject Headings and Subject Headings Used in the Dictionary Catalogs of the Library of Congress. Either of these thesari will provide index terms which can be assigned to the content of an item. If it is desirable to provide in the index, grouping of all items on a given subject by form subheadings can be used (Birds -- Audio Recordings; Indians -- Motion Picture Film). The main index file for the media center should contain all entries for all materials -- print and non-print--for maximum ease of retrieval. This is more important than intershelving and more important than classification.

The depth of indexing media can extend to several descriptors for a single 2 2 slide to indicate production date, country, and art form style, as well as depicted information. Two special indexing schemes dealing with slides have been designed by Robert Diamond and Wendell Simons. Motion picture film can be analysed scene by scene or frame by frame in essence reducing the film to a set of slides. Books can likewise by indexed chapter by chapter especially in the case of collected readings. The decision to be made is how many entry points should be provided for each item.

You might have noted that some of the descriptors mentioned in the previous paragraph and earlier in this paper are not subject-related. Yet descriptors such as date of production or grade level are sometimes desireable. The concept of "added entries" used with print materials can be extended to cover these categories of descriptors. An added entry is an additional entry point to the location of a document under a specific name or place. Added entries are used in printed materials for additional authors, or associations which were involved with the publication.

The file organization method used would generally be a catalog card file or printed list. In the Instructional Media Center highly specialized retrieval is not usually a requirement and the collections are usually of a medium size. For these reasons edge notch and peek-a-boo indexing are not recommended. If easy access to a computer is available, data base indexes could be considered.

In addition to the main index additional supplemental indexes can be created for each form of for special subjects. These can take the form of additional card files or printed lists. It is also possible to set up KWIC indexes for subsets of the collection. Some commercially-produced media indexes include Westinghouse Learning Corporation Learning Directory ; the National Center for Educational Media Index to 16mm Educational Films; and the Library of Congress National Union Catalog - Motion Pictures and Filmstrips.

SUMMARY

Media Indexing should be compatible with all other indexing done in the Instructional Media Center, as well as the same list of terms, with possible additions, should be used and the same file organization. Sears or Library of Congress lists of subject headings would in most cases meet the needs of any center and a total interfiled index (Catalog card file) would provide access.

BIBLIOGRAPHY

Adams, William Mansfield. "A comparison of some machine-produced indexes." (In Studies in indexing and cataloging, p. 425-470) Detroit, Management InformatonServices, 1970.

Artandi, Susan. "Automatic book indexing by computer." American Documentation. 15:250-57 (Oct. 1964)

Association for Educational Communications and Technology. "Standards for cataloging nonprint materials." 3d ed. Washington, DC, 1972.

Borko, Harold. "A research plan for evaluating the effectiveness of various indexing systems." Santa Monica, Calif., System Development Corporation, 1961.

Bottle, Robert and Preibish, Cynthia. "The proposed KWIC index for psychology: an experimental test of its effectiveness." Journal of the American Society for Information Sciences. 21:427-28 (Nov.-Dec. 1970)

Bradley, John G. "Cataloging and indexing motion picture film." American Archivist. 8:165-69 (July 1945)

Cleverdon, Cyril; Mills, Jack, and Keen, Michael. "Factors determining the performance of indexing systems." (In Studies in indexing and cataloging, p. 1-424) Detroit, Management Information Services, 1970.

Collison, Robert L. W. "Indexes and indexing; a guide to the indexing of books, periodicals, music, recordings, films and other materials, with a feference section and suggestions for further reading." 4th ed. London, Benn; New York, De Graff, 1972.

Cox, Carl T. "The cataloging of nonbook materials: Basic guidelines." Library Resources and Technical Services. 15:472-78 (Fall 1971)

Daily, Jay E. "Organizing nonprint materials; a guide for librarians." New York, Marcel Dekker, 1972.

Diamond, Robert M. "The development of a retrieval system for 35mm slides utilized in art and humanities instruction." Fredonia, NY, State Universtiy College, Instructional Resources Center, 1969.

Feinberg, Hilda. "Title derivative indexing techniques; a comparative study." Metuchen, NJ, Scarecrow Press, 1973.

Finn, James D. and Bolvin, Boyd M. "An information indexing, storage, and retrieval system for documents in the field of instructional technology." Los Angeles, Calif., University of Southern Calif., Instructional Technology and Media Project, 1964.

Fischer, M. "The KWIC index concept; a retrospective view." American Documentation. 17:57-69 (April 1966)

Foskett, D. J. "Classification and indexing in the social sciences." Washington, Butterworths, 1963.

Grove, Pearce S. and Clement, Evelyn G. "Bibliographic control of nonprint media." Chicago, American Library Association, 1972.

Hilgard, Ernest R. "Theories of learning." 2nd ed. New York, Appleton-Century-Crofts, 1956.

Johnson, J. T. and others. "AV cataloging and processing simplified." Raleigh, NC, Audiovisual Catalogers, 1971.

Jonker, Frederick. "Indexing theory, indexing methods and search devices." New York, Scarecrow Press, 1964.

Kay, J. B. and Jameson, A. "A classification and retrieval system for recorded foreign language tapes." International review of applied linguistics in language teaching. 5:193-200 (Nov. 1967)

McNally, Paul Terence. "A guide to the practice of non-book librarianship." Toowoomba, Queensland, 1972.

Metcalfe, John Wallace. "Information indexing and subject cataloging. Alphabetical: Classified: Coordinate: Mechanical." New York, Scarecrow Press, 1957.

"Recent advances in indexing." Drexel library quarterly. 8:113-214 (April 1972)

Rice, Stanley. "Picture retrieval by concept coordination: a self-interpreting file." Special Libraries. 60:627-34 (Dec. 1969)

"Sears list of subject headings." 10th ed. Edited by B. M. Westby. New York, H. W. Wilson, 1972.

Shaffer, Dale F. "The audio-tape collection: a library manual on sources, processing, and organization." Salem, Ohio, 1973.

Sharp, John R. "Information retrieval." New York, Seminar Press, 1970.

Simons, Wendell W. and Tansey, Luraine C. "A slide classification system for the organization and automatic indexing of inteedisciplinary collections of slides and pictures." Santa Cruz, Calif., University of California, 1970.

"Standards for cataloging, coding and scheduling educational media." Washington, DC, Natioonal Education Association, Department of Audiovisual Instruction, 1968.

Stevens, Mary Elizabeth. "Automatic indexing; a state of the art report." Washington, DC, Government Printing Office, 1970.

"Subject retrieval in the seventies." (Proceedings of an international symposium held at the Center of Adult Education, University of Maryland, May 14 and 15, 1971) Edited by Hans (Hanan) Wellisch and Thomas D. Wilson. Westport, Conn., Greenwood Publishing Co., 1972.

Sunder, Mary Jane. "Organization of recorded sound." Library Resources and Technical Services. 13:93-98 (Winter 1969)

UNESCO. "International rules for the cataloguing of educational, scientific and cultural films and filmstrips on 3" x 5" (7.5 cm x 12.5 cm) cards." Preliminary edition. Paris, 1956.

Weihs, Jean Riddle; Lewis, Shirley and MacDonald, Janet. "Nonbook materials; the organization of integrated collections." 1st ed. Ottawa, Canadian Library Association, 1973.

Wilson, Thomas D. "An introduction to chain indexing." Hamden, Conn., Linnet Books, 1971.

SELECTED LIST OF THESAURI AND MEDIA INDEXES.

Croghan, Antony. "A faceted classification for and an essay on the literature of the performing arts." London, 1968.

Dailey, Rebecca. "Thesaurus for special education instructional materials." Arlington, Va., Council for exceptional children, 1971.

Dane, William J. "The picture collection; subject headings." 6th ed. New York, Shoestring Press, 1968.

"Educational Media Index." v.1-14. New York, Educational Media Council, 1964.

Educational Resources Information Center. "Instructional Technology subject matter descriptors: a subset of the ERIC thesaurus." Stanford, Calif., ERIC Clearinghouse on Educational Media and Technology, 1971.

Edueational Resources Information Center. "Thesaurus of ERIC descriptors, with a special chapter on the role and function of the thesaurus in education, by F.L. Goodman." New York, CCM Information Corp., 1972.

Engineers Joint Council. "Thesaurus of engineering and scientific terms." 1st ed. New York, Engineers Joint Council, 1967.

"National Cneter for Audio Tapes Catalog, 1970-72." Boulder, Colo.

National Information Center for Educational Media. "Index to 16mm Educational films." 2nd ed. New York, R.R.Bowker, 1969.

New York Times. "Thesaurus of descriptors." 2nd ed. New York, 1969.

Schneider, Erwin H. and Cady, Henry L. "An experimental thesaurus of index terms for documents in music." Columbus, Ohio, Ohio State University, 1966.

Schultz, Claire K. "Information science thesaurus." Philadelphia, Drexel Institute of Technology, 1966.

"Sears list of subject headings." 10th ed. Edited by B.M. Westby. New York, H.W. Wilson, 1972.

"SHE; subject headings for engineering." New York, Engineering Index, 1970.

"Telecourse catalog, 1967." Bloomington, Ind., National Center for school and college television.

U. S. Library of Congress. "Subject headings used in the dictionary catalog of the Library of Congress." 7th ed. Edited by Marguerite V. Quattlebaum. Washington, DC, Library of Congress, 1966.

U. S. National Institute of Health. "Medical and health related sciences thesaurus. 2d ed. Washington, U.S. Public Health Service, 1970.

Westinghouse Learning Corporation. "Learning directory, 1970-71." v. 1-7. New York, 1970.

Williams, Bernard James Stiles. "Thesaurus of microform terms." Hatsfield (Herts.), NationalRerographic Centre for documentation, 1969.

York, E. G. "Subject headings for theological librarianship." Drexel Library Quarterly. 6:53-75 (Jan. 1970)